

# ▲ The Hidden Risk of AI:

## Are Your Unpublished Manuscripts Safe from Plagiarism and Data Leaks?

By Sati Shankar; "ORCID": "<https://orcid.org/0000-0003-4638-1745>", with advisory content by ChatGPT (OpenAI)

### Introduction

In the digital age, where Artificial Intelligence (AI) tools are rapidly becoming part of our writing, research, and creative processes, an alarming issue remains hidden from public discourse:

**Are your unpublished manuscripts being accessed, stored, or even used for training AI systems without your consent?**

This question isn't just academic. It touches on the very core of **Intellectual Property Rights (IPR)**, academic originality, and **creative ownership**. If mishandled, this could be **as dangerous as financial fraud** — threatening the careers, livelihoods, and reputations of scholars, researchers, students, and creators across the world.

### 1. The Growing Dependency on AI-Integrated Writing Tools

From Microsoft Word and Google Docs to Grammarly, ChatGPT, Google Gemini, and Microsoft Copilot — the creative world has begun to rely heavily on **AI-integrated platforms**.

These tools are marketed as productivity boosters. However, most users are unaware that:

- The **default settings** of many AI tools allow them to **store, analyze**, and sometimes **reuse your input data**.
- The **Terms of Service** often contain **ambiguous clauses** that give companies the right to process your data for "service improvement" or "training future models."

This means if you write your manuscript, thesis, or book on these platforms, **you may inadvertently be allowing AI systems to learn from your original content** — even before you have a chance to publish it.

### 2. Can This Lead to Plagiarism? A Serious Risk

**Yes**, and it's not just theoretical.

Imagine this scenario:

You upload a 10,000-word unpublished academic paper to a cloud-based AI editor. Unknown to you, the content is stored in logs or used for training. A few months later, an AI writing tool generates text for another user that reflects your original phrasing, structure, or insights — without credit. You submit your paper to a journal... only to be flagged for "plagiarism" because similar text exists online.

This would be a **catastrophic breach of intellectual authorship**, and sadly, it is now entirely plausible.

### 3. What the Big Tech Companies Say (and Don't Say)

Company	Uses Your Content for AI Training by Default?	Can You Opt Out?
<b>Google</b> (Docs, Gemini)	Yes, unless you adjust your settings.	Yes (settings buried deep).
<b>Microsoft</b> (Word, Copilot)	Says no without consent, but "processing for improvement" may include telemetry.	Yes (enterprise settings safer).
<b>OpenAI</b> (ChatGPT)	Free version may log chats.	Yes, by turning off "chat history".
<b>Anthropic, Meta, etc.</b>	Generally collect inputs unless specified.	Opt-out options limited or vague.

These platforms often claim anonymization, but **even anonymized data can reappear in generated outputs** if the training dataset is rich and contextually narrow (as in academic writing).

### 4. The Ethical and Legal Grey Zone

Intellectual Property Law is still **catching up** with the reality of AI. Most global IPR laws **do not clearly address ownership of AI-generated content or responsibility for AI-trained plagiarism**.

So far:

- There is **no legal guarantee** that your AI-used manuscript won't be indirectly reused.
- Even if you prove AI-induced plagiarism, it's nearly **impossible to identify the responsible party**.

This is particularly dangerous for:

- **Academics and PhD students** working on original theses.
- **Independent authors and researchers** without institutional protections.
- **Scientific manuscripts, patentable work, and cultural or historical research** vulnerable to content mining.

### 5. Steps You Can Take to Protect Yourself

#### ☒ DO THIS:

#### 1. Use Offline Tools for Drafting:

Draft your manuscripts in tools like **LibreOffice, LaTeX, Obsidian, or Markdown editors** that don't connect to the cloud automatically.

#### 2. Encrypt Files Before Uploading:

If you must store documents online, use **end-to-end encrypted services** like **Tresorit, Proton Drive, or Cryptomator**.

**3. Turn Off AI Training Settings:**

For tools like ChatGPT, toggle off **chat history**. On Google or Microsoft, review your **account's privacy and data settings** carefully.

**4. Avoid Copying Full Manuscripts into AI Tools:**

Never paste entire chapters, research findings, or unique intellectual work into AI tools. Use paraphrased queries instead if necessary.

**5. Use “Zero-Knowledge” Cloud Storage:**

Platforms like **Sync.com**, **SpiderOak**, or **Internxt** don't have access to your encryption keys.

## **6. Public Awareness is Critical**

This is not just a private concern — it is a public good.

Much like **financial literacy**, **IPR in the AI age** must become a mandatory conversation in:

- Academic institutions
- Publishing houses
- Research foundations
- Media and journalism schools

We need **global regulations** that make it **illegal to use user data for AI training** without **explicit, informed, revocable consent**.

## **Conclusion**

We are at a pivotal moment in technological history. While AI offers incredible possibilities, it also introduces **a silent but serious threat** to the integrity of creative and academic work.

If manuscripts — the lifeblood of scholarship — are vulnerable to invisible extraction, training, and echoing, then the future of authorship and originality stands at risk.

**Let us act today, to protect what we will value tomorrow: authorship, originality, and intellectual honesty.**

## **📌 Attribution:**

This content was prepared by Sati Shankar; "ORCID": "<https://orcid.org/0000-0003-4638-1745>", with AI advisory from ChatGPT (OpenAI). *You are free to reproduce and circulate it for educational, ethical, and public-awareness purposes with due credit.*

*26<sup>th</sup> of April, 2025, New Delhi, 110001, Bharat*